

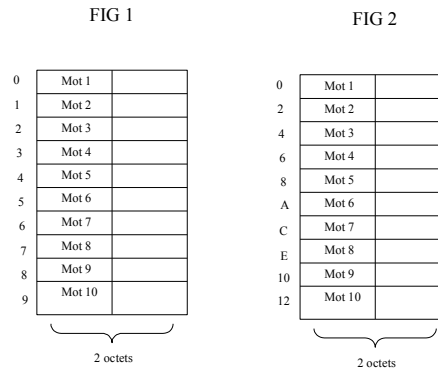
EXERCICES DIRIGES

PRINCIPES DES MEMOIRES CACHES

Rappels de cours

A. Généralités sur la mémoire centrale

- Les mémoires centrales se divisent en deux grandes catégories :
 - Les mémoires vives appelées **RAM** (Random Access Memory) qui se déclinent en DRAM, SRAM, VRAM, etc.... Ces mémoires peuvent être lues et écrites à volonté. Elle sont **volatiles**.
 - Les mémoires mortes appelées **ROM** (Read Only Memory) qui se déclinent en PROM, EPROM, Flash... Ce sont des mémoires à lecture seules qui ne sont pas volatiles. La ROM sur un PC stocke le bootstrap c'est à dire le programme d'amorçage du PC.
- Une mémoire est organisée en **cellules ou mots mémoire**. Un mot mémoire est un ensemble d'octets (1, 2, 4, 8 octets) et est repéré par une **adresse**. Dans une mémoire à adressage par mot, seul les mots mémoire ont une adresse (fig 1). Dans une mémoire à adressage par octets, tous les octets ont une adresse ; l'adresse d'un mot étant alors l'adresse de son premier octets (fig 2).
- La mémoire centrale est reliée au processeur par l'intermédiaire d'un **bus**, qui se décompose en bus de données, bus d'adresse et bus de commande.
- La mémoire centrale demande un minimum de temps pour délivrer l'information lue sur le bus de données. Ce temps est souvent plus grand que le temps de cycle du processeur. Par exemple, avec un processeur à 50Mhz donc de période égale à 20 ns, une mémoire centrale avec un temps d'accès à 30 ns demandera deux cycles horloge processeur pour délivrer son information..



B. Principe de la hiérarchie de mémoire

Sur des critères de coût et de performances, on établit une hiérarchie de mémoire qui vise à donner l'illusion d'une mémoire de capacité illimitée et dotée d'un temps d'accès compatible avec la vitesse du processeur.

Lorsque le processeur désire accéder à une information, il va chercher cette dernière dans l'unité de mémorisation la plus proche et donc la plus rapide. Si l'information est trouvée, on parle de **succès**. Au contraire, si l'information n'est pas trouvée, il y a **défaut** : le processeur effectue un accès au niveau de mémoire inférieure pour y chercher l'information et s'il la trouve, la recopie dans le niveau supérieur.

Le principe de hiérarchie mémoire s'appuie sur le principe de localité :

- **localité temporelle** (localité dans le temps) : si un élément est référencé à un instant t , il y a beaucoup de chance qu'il soit de nouveau référencé à des instants t' proches de t . La localité temporelle guide la recopie d'un élément d'un niveau de mémoire vers le niveau supérieur.
- **localité spatiale** (localité dans l'espace) : si un élément est référencé à un instant t , il y a beaucoup de chance que les éléments dont les adresses soient voisines soient également référencés à des instants t' proches de t . La localité temporelle guide la recopie non pas du seul élément référencé mais d'un bloc d'éléments d'un niveau de mémoire vers le niveau supérieur.

Le cache désigne généralement le niveau de mémoire situé entre le processeur et la mémoire centrale. Il existe trois types d'organisation de cache : les **caches à correspondance directe**, les **caches associatifs** et les **caches mixtes**.

Exercices

Exercice 1

1.1) Soit une mémoire de mots de 32 bits, adressée avec des adresses de 32 bits. La mémoire est adressable par octets. Le cache à correspondance directe contient 4 Ko de données utiles. Une entrée du cache contient un bloc de 1 mot mémoire. La politique d'écriture du cache est une politique de type immédiat.

Calculez la taille réelle du cache.

1.2) Soit une mémoire centrale, adressable par octets, dont la capacité est égale à 256 Mmots de 128 bits. On désire réaliser une mémoire-cache pour améliorer les performances du processeur disposant de cette mémoire centrale en lui ajoutant un dispositif de mémoire cache travaillant par blocs de 4 mots de 128 bits et de capacité utile égale à 512 Ko.

Quelle est la taille réelle de ce cache à correspondance directe, sachant que la politique d'écriture mise en œuvre est de type immédiat ?

Exercice 2

Une mémoire cache à correspondance directe est organisée autour d'une mémoire utile qui contient les données (chaque ligne à une longueur de 16 octets et il y a 16 lignes) et d'un répertoire de 16 lignes (chaque ligne comprend un bit de validation et d'une clé permettant d'identifier la ligne).

2.1) Sachant qu'une adresse mémoire est sur 16bits, calculer la taille des étiquette, index et offset qui composent une adresse.

2.2) On suppose que le cache est initialement vide.

Expliquer si les instructions suivantes mènent à des succès ou des défauts de caches. En cas de défaut sur le cache, le mot mémoire correspondant à l'adresse indiquée dans l'instruction est chargé dans le cache. Indiquez à chaque fois quelle entrée du cache est concernée.

NB : les adresses utilisées dans les instructions (par exemple 0400 dans la première instruction) sont exprimées en base 16

```
LOAD D R1 0400 (Le registre R1 est chargé avec le mot mémoire d'adresse  $(0400)_{16}$ )
LOAD D R1 0832
LOAD D R1 0408
LOAD D R1 0503
LOAD D R1 0836
LOAD D R1 0400
```

Exercice 3

On considère un cache associatif composé de 3 entrées de 4 mots de 32 bits. Les adresses mémoire sont sur 8 bits.

a/ Quelle est la taille de l'étiquette ?

b/ Soit la suite de références suivantes, qui correspondent aux accès mémoire demandés par le processeur, en terme d'adresses d'octets, dans le temps. Les adresses sont données en hexadécimal (base 16)

temps	0	1	2	3	4	5	6	7	8
adresse	0F	1F	3A	0D	1E	44	0B	32	17

Donnez l'évolution des 4 entrées du répertoire du cache et notez les défauts dans les deux cas suivants

1/ la politique de remplacement est FIFO

2/ la politique de remplacement est LRU.